Nayak, Richi and Seow, Lawrence (2004) *Knowledge discovery in mobile business data*, in Shi, Nan Si, Eds. *Mobile Commerce Applications*, chapter 6. Idea Group Publishing (IGI Global).

IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA **ITB9796**
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

Chapter 6

Knowledge Discovery in Mobile Business Data

Richi Nayak, Queensland University of Technology, Australia

Lawrence Seow, Queensland University of Technology, Australia

ABSTRACT

The increasing number of mobile device users is creating a huge amount of useful data for the providers. These data are valuable and can help a business with further developments and strategies if turned into knowledge with the use of data mining. The mindful use of data mining allows organisations to increase customer satisfaction, to determine new consumer groups for marketing purposes, to detect fraudulent activities, and to find future usage of mobile technology. This chapter explores the examples of usage and the process of data mining in the m-business domain. Some of the forthcoming problems to apply data mining in the m-business domain and their possible solutions are also discussed.

INTRODUCTION

Research and practices in mobile (m-) business and mobile (m-) commerce have recently seen an exponential growth. The applications and services that were envisioned for the mobile business marketplace are becoming a reality today (Davis, 2000; Leisen, 2000; McDonough, 2002; Purba, 2002; Tveit &

Tveit, 2002). Mobile activities such as communicating with colleagues via e-mail, booking and purchasing relevant tickets, receiving product information via SMS alerts, and transmitting customer orders with a wireless PDA are typical applications of m-business. M-business applications are not just focused on the consumer use but can also be used on an enterprise level. Using m-business applications, enterprises are able to operate their business more effectively, have a greater level of customer satisfaction, and generate additional revenue (Wireless Business Intelligence, 2002). This means that mobile solutions can influence the way companies maintain their operations, organise employees, monitor inventory levels, and provide on-site solutions for customers.

In its simplest term, m-business is "dynamic," which allows users to access information and perform transactions and other operations from anywhere at anytime via wireless networks. Consequently mobile business applications are generating a large volume of data through various information sourcing and transactions (Magic Software, 2002; Whatis.com, 2002). The m-business techniques and applications advance together with the expanding amount and complexity of data. M-business applications require the monitoring and mining of time-critical data to make sound financial or organisational decisions. The overwhelming need to get benefits from the generated data has provided an increasing opportunity to analyse the data.

Data mining (DM), or knowledge discovery in databases, is the extraction of interesting, meaningful, implicit, previously unknown, valid, and actionable information from a pool of data sources (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). This valuable and real-time information inferred from the data can be used for decision making. This potentially useful information allows maintaining a competitive edge in our present environment. For example, the rapidly increasing sale of mobile phones and PDAs has resulted in an increased number of service providers. The DM technology can help providers to develop services and sales strategies for future benefits.

With the usage of m-business applications increasing every year, it is imperative to understand the concepts and techniques required for mining vital information for organisational needs. The focus of this chapter is to discuss the importance of data mining within the field of mobile technology. This chapter introduces the state-of-the-art data mining technology in the mobile and distributed environment created by mobile commerce applications. Based on the existing and future work of data mining, this chapter collates some data mining applications within the m-business field. There exists a significant number of challenges to access and analyse such types of data. This chapter addresses these data mining challenges and some strategies that should be implemented for better use of m-business data. This chapter also includes some existing work related to data mining applications to m-business. Although much research has been done in both individual areas, literatures relating the two are scarce.

M-BUSINESS ENVIRONMENT

M-business can be defined as the use of mobile technology in the exchange of goods, services, information, and knowledge. M-commerce, on the other hand, is the execution of transactions done on mobile equipment via mobile networks, which may be wireless or switched public networks. M-commerce represents a subset of all electronic business (e-business) transactions, both in the business-to-consumers and business-to-business areas (Whatis.com, 2002). M-business provides services that a wire-connected e-business can not provide. For example, in Finland, people can buy a soda from a vending machine via a mobile phone by dialling a special code on the mobile phone, with the cost being automatically deducted from the customer's bank account. In Frankfurt, Germany, people can locate parking spaces by dialling a number on the mobile phone. The device displays the location of the nearest empty spot, and the cost of the meter is charged on the user's next phone bill. The basic value chain model of m-business consists of two main areas: (1) content and (2) infrastructure and services (Barnes, 2002). The area of content consists of (1) content creation, (2) content packaging, such as formatting, editing, customising, and combining, and (3) market making, such as content and service selection, etc. The area of infrastructure and services includes mobile transport, mobile services and delivery support, and mobile interface and applications, described in the following sections.

Mobile Transport

The current main network technology is the global system for mobile communication (GSM; Barnes, 2002). There are several existing network technologies such as personal communications services (PCS), personal digital cellular (PDC), high-speed circuit-switched data (HSCSD), general packet radio services (GPRS), international mobile telecommunications (IMT2000), etc. (Barnes). Currently, GSM accounts for approximately 71% of the total digital wireless market (Curran & Craig, 2001). The basic architecture of the GSM network is divided into three broad sections (Scourias, 1996).

Mobile station: This section of GSM consists of the mobile equipment and a subscriber identity module (SIM) card.

Base station subsystem: This section is composed of two parts that communicate across a standardized Abis interface (which refers to the interface between the base transceiver station (BSC) and the base station controller (BTS)) that allows the control of radio equipment and radio frequency allocations in the BTS.

Network subsystem: The central component of this section, called the mobile services switching center (MSC), acts like a normal switching node of the PSTN (public switched telephone network) or ISDN (integrated service digital network). MSC provides added functionality for mobile subscribers, such as registration, authentication, location updating, handovers, and call routing to a roaming subscriber.

Mobile Service

Currently, the unique feature derived from the technology of GSM that differentiates it from older analog systems is the availability of short message service (SMS; Curran & Craig, 2001). SMS is a "*bidirectional service for short alphanumeric messages*" (Scourias, 1996) of up to 160 characters as well as non-text-based messages in binary format. Messages are sent between handsets via an SMS center in a store-and-forward system. The advantage of SMS compared with paging systems is the feature of having a message delivery confirmation message. Another added advantage of SMS is the ability to send and receive GSM voice, data, and fax calls simultaneously, using the signalling path to travel over and above the radio channel. Messages can also be stored in the SIM card for later retrieval (Barnes, 2002; Scourias, 1996). There are several other mobile service technologies such as multimedia message service (MMS), cell broadcast (CB), SIM application toolkit (SAT), wireless application protocol (WAP), PDA Web clipping, etc.

Mobile Interface and Applications

The development and integration of application interfaces are becoming critical to conduct mobile business due to the different nature of communications over the present technology of mobile devices compared with standard personal computers (PCs). There have been two main platforms provided by technology platform vendors (Barnes, 2002).

Microbrowser: This is a browser with reduced functionality as compared to those used on the standard PC, such as Netscape and Internet Explorer, to adhere to the present technologies and capabilities of mobile devices. Some of the market's microbrowser companies are: Openwave mobile browser and Nokia.

Operating System (OS): The OS market mainly for personal digital assistants is dominated by Microsoft, Symbian (a consortium comprising Motorola, Ericsson, Nokia, Psion, and Matsuchita), and 3Com (which is now collaborating with Symbian).

Location-Specific Applications

There are numerous location-specific technologies that are being developed such as Bluetooth, enhanced observational time difference, time of arrival, cell of origin, global positioning system (GPS), assisted GPS, etc. (Barnes, 2002).

These technologies allow conducting location-specific m-business, which is sometimes referred as positioning (p-) business. Employing the knowledge of knowing the location of the user to data mining techniques will definitely be the next stepping stone to the further development of m-business (Barnes, 2002; Deri, 2000).

PROCESS OF KNOWLEDGE DISCOVERY

The continuous explosion of data has prompted the development of the process of data mining (DM), or knowledge discovery in databases (KDD), which derives concrete and concise information from data. DM is defined as an interactive, iterative, nontrivial process of deriving valid, interesting, accurate, potentially useful, and ultimately comprehensible structures from data (Fayyad et al., 1996; Freitas, 2001). The data mining process is usually divided into many subtasks, as illustrated in Figure 1. Following is the description of each step involved in the DM process applied to m-business data.

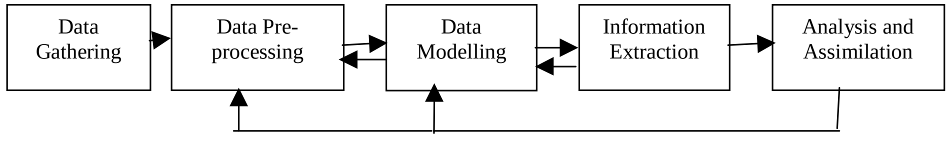
Ascertain Business Objective

Prior to commencing any data mining process, it is important that the businesses clearly identify and define their goals, objectives, limitations, as well as challenges with regards to their operations and economic and financial situation. It is essential to choose the important business areas that require DM to be applied to measure and predict possible future outcomes. Performing data mining without a clear objective will most likely result in additional cost incurred that reaps no benefits (Cabena, Hadjinian, Stadler, Verhees, & Zanasi, 1997).

Data Gathering

This is the initial phase where data from multiple sources are collated according to the business objectives. Since each source of data will be sending data in different formats, data is merged together into a common set of data formats. This is an important task in m-business data since massive amounts of data are gathered from various user transactions and browsing.

Figure 1: The data mining process



Data Preprocessing

A prerequisite for successful data mining is having clean and well-understood data. Due to the fact that data is initially derived from multiple sources, the possibility of having incomplete, noisy, and inconsistent data from the initial data merger is relatively high. In order to perform the data mining process effectively and efficiently, it is essential to apply a set of preprocessing techniques to improve the data quality. This phase includes the following three main steps.

Data cleaning: This step is to rectify the issues of missing values, incorrect attribute values (noisy data), and inconsistent data within the datasets. The clean data is ascertained to be accurate and complete.

Data transformation: In order to apply data mining effectively, the cleaned dataset is transformed to a standard format. The operations of transforming the cleaned data include normalization, aggregation, soothing generalization, and attribute construction.

Data reduction: In order to perform data mining efficiently, the goal is to obtain a reduced volume of the dataset that is representative of the overall cleaned data. This dataset should yield better DM results with faster processing performance. This "ideal" dataset is derived from applying data reduction techniques. This includes removing irrelevant attributes that "confuse" the data mining procedure. For example, attributes having a unique value for each tuple in the database record should be removed as they have no predictive power and will not be able to perform any generalization.

Data Mining

This phase is concerned with the analysis of data by utilising mining techniques to derive hidden and unexpected patterns and relationships from the set of cleaned data. The task is to select a model that fits the end users' needs. There are four main operations associated with data mining techniques.

Predictive modelling: If the goal is to predict future needs, the best model is predictive modelling. This model predicts future events based on previous data by recognising the distinct characteristics of the dataset. A decision tree is developed or a neural network is derived through the analysis of data characteristics. These models are developed over training and testing phases. There are two basic specializations within predictive modelling—**classification** and **value prediction**—according to the type of variable(s) it is inferring to. Classification establishes a specific class to each record in the dataset. There is a finite set of classes that all the data is classified to. Value prediction, or regression, develops a model that is able to estimate a continuous numeric value associated with a particular record.

Clustering: The purpose of this operation is to allow the partitioning of data into segments, or clusters. Data within a segment have high similarity but between segments will have low similarity. This will allow differentiation of homogeneous records (that have close proximity) and heterogeneous records (that are not similar to each other). This operation can be applied using either demographic or neural clustering methods. These two methods are distinguished by (1) the types of input data allowed, (2) the methods of calculating the similarity between records, and (3) the way in which the resulting signals are organised for analysis.

Link analysis: This operation is concerned with establishing links between individual or sets of records. There are three specializations of link analysis: association discovery, sequential pattern discovery, and similar time sequence discovery. The underlying principle of link analysis is to determine the confidence level of the association. The association rule Aà B with the confidence level of 60% means that if a customer has an item A, there is 60% chance that the customer wants to have the item B. This provides the opportunity for businesses to attempt to try and understand users' preferences and needs.

Deviation detection: This operation is concerned with detecting any anomalies, or unusual activities, within a dataset using summarization and graphical representation. The result of this operation brings attention to the business when there are changes deemed material to the well-being of the business.

The data mining phase includes the selection of data mining operations and then appropriate solving techniques.

Knowledge Interpretation

This final phase involves the analysis of the mined results. When the mined results are determined insufficient, an iterative process of performing preprocessing and data mining begins until adequate and useful information has been obtained. Once useful patterns and information have been mined, the post-processing phase ensures the assimilation of knowledge. This involves two main challenges: (1) presenting the new findings in an understandable and convincing way to businesses and (2) formulating ways to thoroughly exploit the new information to benefit the business.

Data Mining Versus Traditional Querying and Reporting Tools

Traditionally, querying and reporting tools of relational database management systems are used to identify specific trends and patterns within the huge amounts of daily transaction data created by m-business. The users of these tools know what kind of information is to be accessed and analysed. Data mining, on the other hand, allows the user to source out unknown facts, i.e., information that is hidden behind the data. This type of data extraction allows business users to seek out new business opportunities and previously unknown data patterns. Another disadvantage of using traditional database queries and reporting tools is the limitations of the output. It is possible to typify questions such as “which mobile service is the most used for users between 20 and 25 years of age.” Data mining enables users to pose more complex queries. For example, DM can predict the estimated sales for the years of interest according to the previous year’s data. Performing a traditional SQL query that provides the same output as data mining is very computationally expensive. Also, time dimension management is not well supported in a relational model.

DATA MINING OPPORTUNITIES IN THE M-BUSINESS DOMAIN

The consumer market is adapting to the thought of using m-business in their daily life. The convenience and opportunities that data mining can provide to m-business are nearly unlimited. There are many areas, including customer relationship management (CRM) and marketing, where currently many commercial e-business application developments and researches are conducted actively (Bowi, 2002; Magic Software, 2002; MobileIN.com, 2002; Regisoft, 2002; Sarwar, Karypis, Konstan, & Riedl, 2000). The possibilities of mining the mobile data sector include nothing less than guiding a nascent industry to fulfil its promise. This section attempts to summarise some applications that data mining can assist in the growing field of m-business.

Taking Advantage of Location Information

Geographic aspects like location and temporal aspects are very interesting in the mobile environment. The GPS (global positioning system) mobile technology has allowed for the identification of the location of the users at a time (Cousins & Varshney, 2001; Duri, Cole, Munson, & Christensen, 2001). By mining location information, we can see the subscribers’ behaviour over time. For example, knowing the locations that a person frequents will allow the prediction of a user’s daily life. If a person is an office worker, it can be noted that most of the time in a week, this person should be present in the work area and at home. Then we can offer location-based services that meet their needs.

Since a phone company already knows some personal and demographic information for all subscribers. With knowing the location-related information, we can generate a dataset with attributes such as subscriber ID, age, gender, marital status, employment, income, place of location, time of location, etc. The *clustering data mining technique* can be used to group subscribers according to their similar interests and can predict what the personality of the person is. Therefore appropriate location-based services can be offered to specific groups of subscribers. For example, if a person is most often sighted in supermarkets and department stores and at home and is seen shuttling between sales events, then this person can be classified as a possible homemaker who is generally interested in events that offer a discounted purchase. Information about sales events near to their place of residence or visit can be sent to a group of subscribers or to an individual subscriber according to their classification. Another example is if a person is often present in the locations of educational institute, pubs, and concert events and the age group is between 20 to 30 years of age, the person can be inferred as an outgoing person. Likewise, information about music events can be sent to subscribers who have an interest in music, or information about nearby takeaways can be sent to office workers working in late hours. Later on it can be analysed whether subscribers are taking advantage of the information suggested by operators, so that operators can take decisions and actions to improve services or increase revenue.

Knowing the location and time of visits for each subscriber, *associative data mining* can also be used to indicate which places a person is most likely to visit in a single trip or in two consecutive trips. This information can be used to suggest a new person to visit the place B if the person is on the place A based on experience of previous visitors.

In terms of a business-to-consumer (B2C; Varshney, 2001) relation, such information will allow business to provide the appropriate marketing information to the specific category of users. For example, users that are categorized as the type most interested in the latest financials news are unlikely to be interested in the information like the latest gardening tools. Thus, a business is better able to identify the needs of users and customize its marketing services to the users. Providing all information to every user, ignoring the user's interest in the type of information given, can result in the user unsubscribing to the services offered by the business.

Issues pertaining to a business-to-business (B2B; Varshney, 2001) relation are also essential to consider. The ability to track the location of employees is ideal to a business to determine the work efficiency of employees. For example, if the worker is sent to a client's location to perform certain duties for the day, the worker's whereabouts are known. If the worker does not go to the client location as directed, that will show the possible work attitude of the worker. Thus, it is ideal to mine the analysis of the worker's work efficiency and attitude to determine the worker who is performing the best and is most suited for the rewards. With the ability to track the location of employees, we do not just have the input attributes, such as employee details, duties performed, and time to finish the duties, but we also have attributes like locations and associated times to perform the duties. We have such information about previous employees too. This previous information can be used as an input dataset in *predictive data mining*, which recognises the distinct characteristics within the dataset due to which a worker is supposed to be working efficiently. Based on this information, a new employee's performance can be measured and suggestions to improve the work efficiency can be provided.

Businesses like courier companies are highly dependent on information regarding the locations of the transported parcels. Having the knowledge of the locations of the parcels allows courier companies to make more informed decisions. For example, if a parcel is to travel between locations A and B, the courier company can do a comparison between the same situation over the last few months using *predictive data mining* and determine if the operation has been efficient or not. If the time to transport a parcel between two locations is much longer for a particular situation, the management can enquire whether it is due to a new staff at a checkpoint that is slowing down the operations. Or is it due to the driver making mistakes, travelling a much further route? Such time-critical issues are vital to a courier company to keep up the standards of their services, which determine the future survival of the company.

Personalization of M-Business Applications

Managing information is becoming one of the biggest issues today. It is important to provide the information that the customer is interested in, referring to the personal life situation and lifestyle. Personalization is particularly useful in areas like information sourcing, whereby data mining can be used to source information that the user specified. Short message service (SMS) is used primarily for simple person-to-person messaging. An increasing number of mobile information services, such as news, stock prices, weather, and notification of e-mail and voice, enables one to aggregate SMS technology and DM technology for providing as one-to-one or one-to-few information services. Information, obtained from analysing the user data about accessing these services using *predictive* and *clustering data mining*, can be used to create personalized advertisements to the customers, delivered by SMS.

Data mining has already been used in e-business to personalize or customize information tailored to the needs of users. M-service providers can also use personalization to provide the products or services that match the needs of individual users. For example, the service provider can use a *clustering data mining* tool to select hotels, either five-star hotels or three-star hotels or hotels in the suburbs or downtown according to the customer preference, based on previous similar travel service experiences. The clustering data mining technique groups customers with similar preferences. When a new customer mentions his/her preferences, a similar preferences cluster is matched, and based on those preferences a recommendation is made.

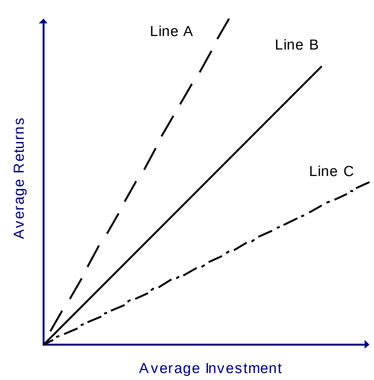
Relevant services can be offered based not only on the personal profile of the device holder but also on the device holder's location and time factor. For example, m-business applications used in the travel industry can assist users to find attractions, hotels, and restaurants of their preference on requested location and time. *Associative data mining* can be used to indicate which places a person is most likely to visit in a single trip or in two consecutive trips, with having inputs such as location and time of visits to attractions for each user. This provides great convenience for users as these services can be used while driving; for example, a suggestion can be given based on the association rule that if the user is on place A then the user should visit place B since 80% of previous visitors have done so.

Due to the limited screen space provided on mobile devices, it is difficult for mobile users to browse the product or service catalogues on the devices. Thus, not only personalization services offered by the vendors can provide great convenience for customers, these services can also assist to retain customers that are critical in today's competitive environment.

Trend Analysis of Costs Versus Benefits

Any m-business providing mobile facilities constantly analyses whether the profits derived from the business are sustainable. Data mining is essential to do a trend analysis of the business over a period, for example, to determine if the progress of the business is of a reasonable expectation. One of the techniques to determine the profitability of the business is the calculation of the return of investment (ROI). ROI is typically calculated as: (average returns / average investment) * 100% (Dickerson, Campsey, & Brigham, 1995). A possible way to analyse the data collected from the profit derived from the business is the use of linear regression (a *value prediction technique* of the DM) to predict the prospect of the business. A graph can be plotted to provide visualization and to allow the analysis of the future and potential of the business. Figure 2 illustrates

Figure 2: Linear regression to determine viability of business



the graph based on average returns verses average investment into the business. Many other business factors can also be considered during regression analysis.

Scenario 1, the best-fitting line derived from a variation of investments appears as Line A. This shows that the investments made into the business have been worth undertaking, as the increase in profit is greater than the increase in investments into the business.

Scenario 2, the best-fitting line appears as Line B. This means that the average return derived is equivalent to the average investment. Thus, it is neither profitable nor detrimental to carry on the business.

Scenario 3, the best-fit line appears as Line C. This means that the cost incurred to provide the mobile marketing service is greater than the revenue generated from the business. This shows that the mobile service provider has to rethink its operational and management plans.

Predicting Customer Buying and Usage Patterns

In the near future, the PDA or the mobile phone will be our wallet, as most services will be payable with a few clicks on our devices. The service providers will have an enormous amount of consumer behaviour information stored in their databases. Patterns in consumer behaviour can be analysed from many angles, and information can be extracted to benefit further business implementations. M-business with DM provides new channels to marketing that can be used to provide a much more direct and to the point advertising to the customers.

Service providers can analyse the data (e.g., by analysing gateway log files and content server log files on WAP) and predict the consumers' buying and usage patterns or understand how mobile subscribers used their wireless services. Using this data, companies can apply data mining to identify customer segments using *clustering data mining* techniques, to distinguish customers' consumption patterns using *deviation detection techniques*, and to predict transaction trends using *associative data mining techniques*. This information then can be used to provide better services to the customers or to attract potential customers.

For example, if customers are buying several services, *predictive data mining* can help the service provider to forecast the users' service needs. Users seem to change their usage and needs; these changes need to be captured and analysed to predict what the users want/need in the future. Inputs to the predictive data mining process will be the details of users (personal, demographic, and geographical) who are using the services and also the details of users who are not using the services. The *predictive data mining* engine will be able to establish rules, based on distinct features, why certain users are subscribing these services.

If customers are using a combination of two or more services, *associative data mining* can help the company to provide better services to subscribers.

Providing inputs such as customerID and the services used by them, *associative data mining* can infer rules such as 75% ofcustomers who use services A and B also use services C and D. This will help service operators to adjust prices on packages to get more customers to use all. These are just a few examples. Essentially, the customer purchasing information is stored on a digital storage medium and this gives the m-businesses new opportunities to increase their markets.

Predicting Future and Better Usage of Mobile Technology

All the patterns between mobile phone brands and usage of available services can be extracted and used to further marketing. Data to analyse in this situation are the amount of mobile phones in the market, how many of these users use the services available, and how much usage, measured in currency, that the average user uses. This data analysis using *classification* or *regression data mining techniques* will be able to predict the trends and patterns of usage of mobile phones and mobile services.

For example, the most popular services bought through m-commerce technology are mobile ringing tones, logos, and screensavers. The most commonly used interfaces for these kinds of transactions are short message service (SMS) and the standard e-commerce interface, the Internet. An example is Nokia's focus on screensavers, logos, and ringing tone availability. This is most likely to be a result of previous research on their users' trends, by capturing the data on users' demands and needs and then analysing users' feedback. This information helped Nokia to develop a new market product where the product is no longer just a mobile phone but also provides extra features like SMS, logos, and additional ringing tones and screensavers (Nokia, 2002). These new features can be bought using the phones. All these transactions between the users and the service providers, and the patterns of customers buying ringing tones, logos, screensavers, SMS pictures, etc. can be analysed for further actions by the use of data mining tools.

A mobile commerce platform should integrate with existing back-end databases and businesses applications to deliver data via all the channels such as WAP, VoxML, TruSync, Bluetooth, or any wireless protocol. Data mining can be used to match which channel is best at a time to deliver the information. Data mining can optimise the amount and format of the content for delivery based on the connection speed of the device requesting the information. Data mining helps to decide what tasks, activities, and transactions are most economical and beneficial to use.

Target Marketing or Direct Marketing

Unlike e-commerce in which the customer accessing the company site is anonymous, it is easy to identify the customer doing m-business and track their behaviour from various sources to build their profiles. This information can be used for marketing purposes. *Clustering data mining techniques* can be used to segment groups of people that appeal to certain products or services. Knowing this information will allow the marketing department to come out with "gimmicks" to target specific customers. Also, *associative data mining techniques* can be used to identify certain items that customers tend to buy together. This information will allow m-businesses to offer discounts on various purchases without bearing any losses, in turn attracting more customers and revenue.

Fraud Detection in M-Business

The effective and subsequent analysis of the types of fraudulent activities in telecommunication systems is one of the applications that data mining can assist in a mobile environment. The dynamic nature of different fraudulent activities and the changes of the normal usage can lead to detection of fraudulent activities through observing behavioural patterns. A data mining system will have plenty of examples of normal usage and some examples of fraud usage. Based on these previous examples, a *predictive data mining* system establishes facts about fraudulent activities. Whenever a change in the normal usage is detected, the system analyses the change and will be able to predict whether the change is a fraud or not.

Churn Management in M-Business

Churn management is a term used in the telecommunication industry to describe the process of ensuring that profitable customers stay with a particular company. *Predictive data mining* techniques can assist in churn management by forecasting whether a given individual is likely to move to another service provider by analysing their usage patterns, and these techniques are able to define the correct actions to keep that profitable customer (Purba, 2002; SAS, 2002).

DATA MINING CHALLENGES AND THEIR SOLUTIONS

With the increasing use of mobile services, it is very likely that mobile devices will use the concept of data mining in the future. In order to apply data mining efficiently in m-business, certain requirements have to be met. Ideally the methods used for mining mobile data should be able to (1) mine different kinds of knowledge in databases, (2) deal with diverse types of data types such as relational, temporal, and spatial types of data, (3) mine information from heterogeneous databases and global information systems, (4) handle noisy and incomplete data, which is mostly the case in the m-business domain, (5) perform the mining tasks efficiently regardless of the size and complexity of the dataset, (6) support interactive mining of knowledge at multiple levels of abstraction, (7) support integration of the discovered knowledge with existing knowledge, and (8) deal with the issues related to applications of discovered knowledge and social impacts such as protection of data security, integrity, and privacy.

The process of obtaining useful information from voluminous records of actual mobile sessions data calls for using powerful, parallel, distributed, scalable, integrated, and incremental data mining tools. The data mining software can be developed as a collection of components that may be based on object technology. By developing data mining modules as a collection of components, one can develop generic tools and then customize them for specialised applications. This section attempts to summarise the requirements and the future issues that need to be addressed when data mining is applied in the mobile sector.

Distributed Environment

In the m-business environment, data can reside in many different geographical locations. Most data mining systems are currently based on centrally located data; data is stored in a single database and the mining techniques are focused on this dataset. As a result of convergence between computation and communication, the new data mining approaches have to be concerned with distributed aspects of computation and information storage. This means that organisations will have to implement decentralised approaches for data storage and decision support. Distributed data mining typically involves local data compression and analysis for minimisation of network traffic as well as the generation of global data models and analysis by combining local data and models (Park & Kargupta, 2002).

A development of integrating data mining applications, data mining systems, and business processes effectively together will guarantee and support the environments of e-business and m-business. In order to conduct data mining in a distributed environment where data is collected from multiple sources, XML is proving to be the most ideal solution to realize such a potential. Every mobile device is able to transmit XML documents that can be read and processed easily, regardless of which platform the mobile device is running on. XML has provided the facilities to perform data exchange on the Web as well as wirelessly between applications or between users and applications in a flexible and extensible representation (Graves, 2002).

However, there are always the common situations of incomplete, noisy, and inconsistent data. Even though the imperfect dataset is cleaned during the preprocessing step, nevertheless, the data are never cleaned 100% perfectly.

The appropriate procedures should be developed to cleanse the received and the integrated imperfect data.

Clickstream Data

At present, the Web environment has catered to electronic commerce (Nayak, 2002). With millions of people accessing the Web every day, it has been possible to gather a large amount of "clickstream data" (Kohavi, 2001) to determine and predict the possible interest of users. But at the current level of technology, mobile devices have a number of limitations. One of these limitations is having a small display screen (Madria, Mohania, Bhowmick, & Bhargava, 2002). As a result, on a WAP phone, the average number of links it has to other Web sites is an average of 5 links, while a standard Web page has an average of 25 links. If a user is to have three clicks on the Web via a WAP phone, there are only 5^3 (= 125) pages that are accessible to the user, compared with the standard Web page having 25^3 (= 15,625) accessible pages (Barnes, 2002). Therefore, users of mobile devices are highly restricted on the Web pages that they can visit. It is quite unlikely that the user will be going to the site that he really wants from the links available. As a result, the usage of data mining to analyse clickstream data collected from users of mobile devices is not going to be accurate. Predicting the user's interest will be difficult. Data mining is therefore limited due to the current mobile Internet service, which is not able to reflect the user's interest sufficiently.

Security and Privacy

With the technology of sending messages to mobile users, it has become possible for users to specify the types of information that they prefer and require. In a mobile network, this is known as creating "dynamic bookmarks" (Duri et al., 2001). As a result, a business is able to provide users with the information and services that they prefer. For example, if the user indicates that his preference is a particular brand of product above a particular price, then it can be analysed that the user may also be interested in another similar brand of the same standard as well. This opens data mining possibilities such as classifying the users based on their reported needs using *predictive mining* and *clustering* and finding correlations between various needs by performing a *link analysis*.

Unfortunately there are also situations when users are not comfortable about declaring any information about them. As a result, preferences indicated by users might not necessarily be correct. Thus, data mining results may have classified the user as a potential person to send information to, but in reality, it could all end up as an added expense to incur cost in conducting the data mining techniques and including the irrelevant people into the mobile service.

There is also a certain level of fear within users with regards to mobile data security. Security is a vital concern in m-business due to the type of communication medium. There is always a potential risk of compromising the integrity, security, and availability of information with the portability of mobile devices (Madria et al., 2002). This exacerbates the possibility of users who do not believe in the security of mobile data to inaccurately declare their personal information and personal preference. Also the vocal type of data transfer mode is not appropriate for applications with confidential data where one could be overheard.

Cost Justification

With the issue that the data mining application is usually computationally expensive, there is always a concern whether the benefits of data mining justify the cost incurred for the process. Firstly, before data mining commences, preprocessing of the data is required to ensure the data is cleaned and all inconsistent or missing values are adequately rectified. A substantial amount of computational power is required to perform this process. Secondly, the cleaned data needs to be transformed to an appropriate format to facilitate the mining process. With the concern of security and privacy of XML documents being transferred wirelessly, XML allows its documents to be complex and tagged with unmeaningful names. So the document is not useful to an unauthorized person without the knowledge of how to transform the document appropriately. However, the more complex the XML document is, the more computational power is needed to process these documents. Thus, there is a difficulty to strike a balance between the security and privacy of data transferred versus the computational cost required to process the "encrypted" documents and also between the cost involved in the data mining process versus its benefit.

Technological Limitation

Although there are a vast number of mobile technologies available, there exist several limitations and constraints of the technologies adversely affecting the performance of data mining in the m-business domain. Some of the limitations with regards to mobile computing include low bandwidth, limited battery power, and unreliable communications that result in frequent disconnections and higher error rates. These factors have resulted in the increase in communication latency, additional cost to retransmit data, time-out delays, error control protocol processing, and short disconnections (Madria et al., 2002). There are also many possibilities of lost connections like when a user moves to areas of high concentrations (e.g., events, concerts, etc.) or interference. These limitations pose significant problems in collecting the data for mining purposes.

Furthermore, technologies like GPS have allowed the location of users to be identified outdoors. The idea of knowing a user's location appears promising in data mining; for example, providing tourists with a map of their current location on their mobile devices (Brown, Chalmers, & MacColl, 2002) or possibly even sending messages to consumers when they bypass a shopping sale. However, present technologies are not mature enough to determine the location of a user indoors (Duri et al., 2001). Although the potential of gathering knowledge from a user's location is very useful in terms of m-business and data mining, this potential is not yet realized until present technologies improve to provide adequate and up-to-standard location-aware applications to function on mobile devices.

At present, only a minority of consumers have mobile devices that allow for wireless Internet connections. This is due to the current technology limitations such as devices with small screens and only allowing the viewing of Internet contents via a scaled-down version of HTML (Hypertext Markup Language)—WML (Wireless Markup Language)—which does not support Internet viewing as good as on a personal computer. Limitations in wireless network bandwidth (ranging from 9.6 to 19.2 Kbps) do not also completely live up to the high expectation of consumers (Kalakota & Robinson, 2002). Thus, technologies in mobile devices are perceived as not stable and reliable as yet.

Data mining over distributed sources would definitely be limited due to the technological limitations of the devices. For example, the present low bandwidth means that the data transfer is slow. This implies that data mining processes have to be delayed until most data transfers have been completed and received. Web serving is already very popular in e-business. In future there will be additional resources for Web access. One of them is the wireless access of Web sites via mobile phones. Unfortunately the bandwidth is a big issue and therefore it is essential to minimize the data volume that is sent through the channels. We have shown earlier in the paper that with the possible DM opportunities in m-business it is possible to overcome some of these limitations.

Data Mining Process and Intelligent Agents

The several steps of the knowledge discovery process can be partly automated using intelligent agents (Russell & Norvig, 1995). Intelligent agents use domain knowledge with embedded simple rules. The use of training data helps to reduce the need for domain experts. A scanning agent goes through the rules and facts and displays the items that have valuable information.

Intelligent agents can help to automate the data selection steps by determining learning parameters, by applying triggers for database updates, and by managing invalid data. The agents can perform automatic sensitivity analysis to detect helpful parameters. As a result the number of domain experts will be reduced, so there is no more need of experts whenever the environment is changing. Data cleansing can be automated using intelligent agents with a rule base. Whenever a record is added or updated in a relational database, the trigger of the agent examines the transaction data. Missing or invalid data can also be cleansed by using the rules in its rule base.

Agents can be used in implementing classification, clustering, summarization, and generalization models that have a learning nature and rules generation. The search for patterns of interest, by using learning and intelligence in classification, clustering, summarization, and generalization, is supported by intelligent agents. Agents can generate newly discovered related information from data by learning the preferences from a profile or from examples and feedback from a user profile. This information can be used to provide confidence in what the agent is predicting. The agents are implemented based on machine learning techniques and data mining techniques such as case-based reasoning, neural networks, association, and induction (Seydim, 1999).

One major advantage of intelligent agents is in their support for online transaction data mining. This automated decision support is called "active data mining" (Moon, Kim, & Kim, 2001). In conclusion intelligent agents are very important in the process of knowledge discovery, especially in distributed environments such as m-businesses, by supporting the discovery process in many stages.

EXISTING WORK

This section discusses some existing data mining works within the m-business domain.

MobiMine—Monitoring the Stock Market From a PDA

MobiMine is an experimental mobile data mining system which enables intelligent monitoring of time-critical financial data from a handheld PDA. The system consists of several modules, with some of the modules utilising data mining techniques. The data mining component of the system employs a novel Fourier-analysis-based approach to efficiently represent, visualize, and communicate decision trees over limited bandwidth wireless networks (Kargupta et al., 2002; Kargupta, Sivakumar, & Ghosh, 2002).

MobiMine is a client-server application. The clients running mobile devices likehandheld PDAs and cell phones monitor a stream of financial data coming through the MobiMine server. The MobiMine server and client apply several advanced data mining techniques to offer the user a variety of different tools to monitor the stock market at any time from anywhere. The server collects stock market data from different sources on the Web and processes it on a regular basis. It employs several data mining techniques to sift through the data. MobiMine makes use of a collection of online mining techniques, including several statistical algorithms, clustering, Bayesian nets, and decision trees. The StockConnection module uses online statistical Fourier-spectrum-based decision trees and Bayesian learning techniques for detecting the interaction among the active stocks and the portfolio. The StockNuggets module applies a collection of different online clustering algorithms for identifying interesting stocks that are influenced by the current active stocks.

Game Usage Mining: Knowledge Discovery in Massive Multiplayer Games

Computer games are still one of the most demanded applications in the electronic environment. The primary purpose of data mining in games is to identify different patterns of behaviour, structure, or content in order to improve the overall game play. With this data, organisations can pinpoint areas that need improvement to increase player satisfaction. In return, this will increase revenue and reputation. Wireless Internet revenue models are either proportional to money per time spent by each player (e.g., WAP over GSM) or money per byte served to the player (e.g., UMTS, I-Mode, and WAP over GPRS; Tveit & Tveit, 2002).

There are three different main types of data mining approaches referring to wireless games (Tveit & Tveit, 2002):

1. Game content mining—discovery of patterns in multimedia or textual content in games (e.g., room layout)
2. Game structure mining—discovery of structural patterns in the form of paths and connections binding the game world together (e.g., hallways between rooms)
3. Game usage mining—discovery of human and virtual player behaviour patterns

CONCLUSION

In today's society, the use of mobile devices is increasing dramatically. The majority of mobile devices, initially used for telecommunications, have now been enhanced to support business needs and growth. The use of mobile devices in businesses has led to the creation of m-business/m-commerce. The increasing number of mobile device users creates a large amount of useful data for service providers. These data are valuable and can help the business with further developments and strategies if they are properly analysed.

The success of an m-business depends on the ability to deliver attractive products or services that are personalized to the individual user at the right location at the right time. These information-intensive services can only be obtained by collecting and analysing the combined demographic, geographic, and temporal information. The challenge for mobile service providers is to manage the overwhelming data that they are accumulating every day and apply data mining tools effectively to transform those data into useful information that can not be seen with traditional reporting techniques and tools. Data mining enables the user to seek out facts by identifying patterns within data. Data mining can give businesses the edge over other businesses by increasing competitiveness, in the form of marketing that is more focused on particular consumer groups or by suggesting the better use of mobile technology.

Existing applications of data mining in regards to m-business include such works as MobiMine, which enables a user to monitor stock prices from a handheld PDA. Applications like this will increase dramatically within the near future to accommodate the need for business expansion and optimisation. An investment in a data warehouse and a data mining tool is costly but can help the m-business to provide the right services to the right people at right time, thus proving to support decision making, increase customer satisfaction, and aid in marketing.

In this chapter, we explored examples of usage and the process of data mining in the m-business domain. We also discussed some of the forthcoming problems in applying data mining in the m-business domain and their possible solutions.

REFERENCES

Barnes, S. J. (2002). The mobile commerce value chain: Analysis and future developments. *International Journal of Information Management*, *22*, 91-108.

Bowi Customer Relationship Management Solutions. (2002). *Glossary*. Retrieved September 22, 2002, from http://www.bowi.de/en/glossar/gloscont.htm

Brown, B., Chalmers, M., & MacColl, I. (2002, September). *Exploring tourism as a collaborative activity* (Tech. Rep. Equator-02-018). Retrieved from http://www.equator.ac.uk/papers/Abstracts/2002-brown-6.html

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1997). *Discovering data mining from concept to implementation*. Prentice Hall PTR.

Cousins, K., & Varshney, U. (2001, July). A product location framework for mobile commerce environment. In *Proceedings of the First International Workshop on Mobile Commerce* (pp. 43-48).Curran, K., & Craig, R. (2001, November). A short message service online application for delivering urgent information to students. In *Proceedings of the First Joint IEI/IEE Symposium on Telecommunication Systems Research*.

Davis, J. (2000, November 16). Peek into the future of mobile shopping. *The Industry Standard*. Retrieved September 22, 2002, from http://www.cnn.com/2000/TECH/computing/11/16/m.commerce.future

Deri, L. (2000, May). Beyond the Web: Mobile WAP-based management. *Journal of Network and Systems Management*.

Dickerson, B., Campsey, B. J., & Brigham, E. F. (1995). *Introduction to financial management*. Dryden Press.

Duri, S., Cole, A., Munson, J., & Christensen, J. (2001, July). An approach to providing a seamless end-user experience for location-aware applications. In *Proceedings of the First International Workshop on Mobile Commerce* (pp. 20-25).

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge discovery and data mining: Towards a unifying framework. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*.

Freitas, A. A. (2001). A survey of evolutionary algorithms for data mining and knowledge discovery**.** In A. Ghosh & S. Tsutsui (Eds.), *Advances in evolutionary computation*. Springer-Verlag.

Graves, M. (2002). *Designing XML databases*. Prentice Hall.

Han, J., & Kamber, M. (2001). *Data mining: Concepts and techniques*. Morgan Kaufmann.

Kalakota, R., & Robinson, M. (2002). *M-business: The race to mobility*. McGraw-Hill.

Kargupta, H., Park, B., Pittie, S., Liu, L., Kushraj, D., & Sarkar, K. (2002, January). MobiMine: Monitoring the stock market from a PDA. *SIGKDD Explorations*, *3*(2), 37-46.

Kargupta, H., Sivakumar, K., & Ghosh, S. (2002). Dependency detection in MobiMine and random matrices. In *Proceedings of the Sixth European Conference on Principles and Practice of Knowledge Discovery in Databases* (pp. 250-262).

Kohavi, R. (2001). Mining e-commerce data: The good, the bad, and the ugly. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

Leisen, B. (2000, August 18). M-commerce: Mobile and multiplying. Wall Street Journal, pp. B1, B4/ Retrieved September 25, 2002, from http://myphlip.pearsoncmg.com/cw/mpviewce.cfm?vceid=1913&vbcid=1714

Madria, S. K., Mohania, M., Bhowmick, S. S., & Bhargava, B. (2002, April). Mobile data and transaction management. *Information Sciences*, *141*(3-4), 279-309.

Magic Software. (2002). *Mobile eBusiness Magic white paper*. Retrieved September, 20, 2002, from http://www.magic-sw.com/

McDonough, D., Jr. (2002, April 22). Digital River gives m-commerce a kick. *Wireless NewsFactor*. Retrieved September 24, 2002, from http://www.wirelessnewsfactor.com/perl/story/17377.htm

MobileIN.com. (2002). *Personalization, mobile communications & intelligent networks and applications*. Retrieved September 22, 2002, from http://www.mobilein.com/personalization.htm

Moon, H., Kim, S. H., & Kim, K. C. (2001, December). Adaptive agent for wired and mobile media: Framework and application to multilingual content. In *Proceedings of the International Conference on Computers in Education*.

Nayak, R. (2002). Data mining for Web-enabled electronic business applications. In S. Nansi & V. K. Murthy (Eds.), *Architectural issues of Web-enabled electronic business* (pp. 129-138). Hershey, PA: Idea Group.

Nokia. (2002). *Nokia: History in brief*. Retrieved September 25, 2002, from http://www.nokia.com/aboutnokia/compinfo/history.html

Park, B., & Kargupta, H. (2002). Distributed data mining: Algorithms, systems, and applications. In N. Ye (Ed.), *Data mining handbook*.

Purba, S. (2002). *New directions in Internet management*. Auerbach.

RegiSoft. (2002). *RegiSoft's mobile marketing*. Retrieved September 22, 2002, from http://www.regisoft.com/

Russell, S., & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Prentice Hall.

Sarwar, B. M., Karypis, G., Konstan, J. A., & Riedl, J. T. (2000). Application of dimensionality reduction in recommender system—A case study. In *ACM WebKDD 2000 Web Mining for E-commerce Workshop*.

SAS. (2002). *Telstra Mobile combats churn with SAS*. Retrieved September 24, 2002, from http://www.sas.com/news/success/telstramobile.html

Scourias, J. (1996). *Overview of GSM: The global system for mobile communications* (Tech. Rep. CS-96-15). Ontario, Canada: University of Waterloo, Department of Computer Science. Retrieved September 13, 2002, from http://ccnga.uwaterloo.ca/~jscouria/GSM/

Seydim, A. Y. (1999). *Intelligent agents: A data mining perspective*. Dallas, TX: Southern Methodist University, Department of Computer Science and Engineering. Retrieved September 26, 2002, from www.engr.smu.edu/~yasemin/agentsdm.pdf

Tveit, A., & Tveit, G. B. (2002). Game usage mining: Information gathering for knowledge discovery in massive multiplayer games. In *Proceedings of the International Conference on Internet Computing* (IC 2002).

Varshney, U. (2001, July). Location management support for mobile commerce applications. In *Proceedings of the First International Workshop on Mobile Commerce* (pp. 1-6).

Whatis.com. (2002). *M-commerce*. Retrieved September 22, 2002, from http://whatis.techtarget.com/definition/0,289893,sid9_gci214590,00.html

Wireless Business Intelligence. (2002). Retrieved September 23, 2002, from http://www.digimine.com/solutions/wirelessbusinessintelligence.asp